Journal of the Korean Data & Information Science Society
2012, **23**(6), 1155–1163

http://dx.doi.org/10.7465/jkdi.2012.23.6.1155
한국데이터정보과학회지

사영을 이용한 제2종 분석[†]

최재성[1]

[1]계명대학교 통계학과

접수 2012년 10월 17일, 수정 2012년 11월 11일, 게재확정 2012년 11월 17일

요약

본 논문은 이원 고정효과모형의 가정하에 실험자료를 분석할 때 고정효과와 관련된 변동량을 구하는 문제를 다루고 있다. 고정효과에 따른 변동량의 계산을 위한 방법으로 제2종 분석을 이용하고 있다. 모형비교의 방식을 이용하여 고정효과에 따른 변동량을 계산하는 제2종 제곱합은 비교하는 모형의 적합에서 주어지는 잔차제곱합의 차에 근거를 두고 있다. 이와는 달리, 본 논문에서는 고정효과와 관련된 모형행렬로의 사영에 의한 변동량의 계산방법을 다루고 있다. 사영에 의한 제2종 제곱합의 계산은 기존의 방법보다 간편하고 효율적임을 보여준다.또한 사영에 의한 제2종 분석으로 변동량의 계산에 있어 균형자료와 달리 불균형자료에 있어서 어떤 점에서 차이가 있는 가를 구체적으로 논의하고 있다.

주요용어: 고정효과모형, 불균형자료, 사영, 제2종 분석.

1. 서론

실험을 통하여 비교하고자 하는 처리들의 효과에 대한 추론은 대개 선형모형의 가정하에서 행해진다. 가정된 선형모형이 고정효과 모형으로 간주될 때 자료의 분석기법 은 일반적으로 최소제곱법에 의한 분산분석법을 이용한다. 분산분석법에 의한 자료의 변동량 분석은 자료가 행렬로 표현될 때 사영에 의한 방법으로 제공될 수 있다.

개체의 반응에 대한 실험자료 또는 관측자료가 다차원 공간상의 벡터로 표현되면 모형의 행렬 표현식으로부터 관측벡터는 모형행렬로 생성되는 벡터 부분공간으로의 사영에 의해 추정되므로 사영의 개념에 근거한 사영행렬의 이용이 가능하다. 사영의 관점에서 행해지는 변동량의 분석은 단순히 행렬의 성질을 이용한 일반적인 분산분석방법과는 두 가지 측면에서 접근방식이 다르다고 볼 수 있다. 하나는 사영에 의한 분석방법은 최소제곱법에 비해 분산구조에 있어 제한적이지 않다. 다른 하나는 사영공간을 나타내는 사영행렬은 모형에서 고려되는 변수들의 좌표축을 나타내는 고유벡터와 축에서의 변이를 나타내는 고유근을 포함하고 있으므로 변동량의 분석에 있어 이들을 다양하게 활용할 수 있다는 점이다.

변동요인에 따른 제곱합을 구하는 방법은 관련된 모수를 포함하는 모형의 적합방식에 따라 구분된다. 제1종 분석은 모형내 모수의 순차적 적합에 따른 잔차제곱합 간의 차를 이용하여 관련된 모수의 변동량을 계산하는 방법이다. 제2종 분석은 관심모수와 관련된 변동량의 계산을 모형비교의 방식을 이용해 분석하는 방법이다. 자료가 균형자료일 때는 모형의 적합방식이 달라도 자료전체의 변동량은 변동요인에 따른 제곱합의 총합과 일치하게 된다. 그러나 자료가 불균형일 때 자료의 전체 변동량을

[†] 본 연구는 2012년도 계명대학교 비사연구기금으로 이루어졌음.
[1] (704-701) 대구광역시 달서구 신당동 1000번지, 계명대학교, 통계학과, 교수. E-mail: jschoi@kmu.ac.kr

나타내는 총제곱합은 변동요인에 따른 제곱합의 총합과 일치하지 않게 된다. 이러한 차이에 대한 근거를 행렬의 대수적 방법으로 입증하는 것이 단순하지 않으므로 사영의 관점에서 논의해 보기로 한다.

변동요인에 따른 제곱합을 제2종 분석에 의해 구하는 경우에 그 값은 모형비교에 따른 잔차제곱합의 차 또는 관심모수의 적합에 따른 감소량으로 계산됨을 의미한다. 즉, 제2종 제곱합은 모형비교를 통해 주어지는 변동량의 차이로 결정되므로 사영의 개념에 근거한 사영분석과는 차이가 있다. 왜냐하면, 사영분석에서는 제2종 제곱합을 사영이 행해진 사영공간에서 사영까지의 거리제곱으로 구하기 때문이다. 따라서 제2종 분석방법에 의한 요인별 제곱합을 사영에 의해 구하기 위해 모형행렬로부터 생성되는 사영공간으로의 사영과 사영행렬을 이용하는 방법을 구체적으로 다루고자 한다.

제1종, 제2종 분석에 의해 구해지는 변동요인들의 제곱합은 동일 자료의 사영에 의한 분석에서도 같은 값으로 얻어진다. 즉, 모형의 적합방식에 따른 변동요인의 제곱합은 사영공간에서의 사영행렬을 이용하여 동일하게 구해질 수 있다. 따라서 사영에 의한 분석방법은 분산분석에 있어서 모형의 적합방식에 제한적이지 않다고 할 수 있다.

실험자료를 분석하기 위한 모형으로 선형모형을 가정하게 될 때, 실험단위의 반응에 영향을 미치는 처리요인들의 수준이 어떻게 선택되는 가에 따라 특정한 유형의 선형모형의 가정하에 자료를 분석하게 된다. Searle (1971)과 Graybill (1976)은 여러 확률모형에 대한 분산성분의 적률추정량과 관련된 성질들을 다루고 있다. 이외에도 처리구조에 반복측정요인이 포함될 때 Galecki (1994)는 공분산구조를 다루고 있고 Choi (2010)은 반복측정의 분할구 실험자료에 대한 모형과 분석방법을 제시하고 있다. Milliken과 Johnson (1984)은 실험계획을 처리구조와 설계구조의 두 구조로 다루고 있다.

본 논문은 처리구조를 나타내는 요인이 모두 고정요인인 경우를 가정하고 고정효과 모형 또는 고정모형의 가정하에 모형내 모수들의 효과와 관련된 제곱합을 구하는 방법에 관심을 두고 있다. 고정효과 모형의 가정하에 실험자료를 분석하는 일반적인 방법은 분산분석법이다. 분산분석에 이용되는 모수 추정방법은 주로 최소제곱법을 사용하게 된다. 모수를 추정하기 위한 최소제곱법의 적용은 관측자료가 다차원 공간상의 관측벡터로 간주될 때 모형행렬로 생성되는 벡터 부분공간으로의 사영을 의미하게 된다. 일반적인 분산분석법에 의해 제공되는 변동요인의 제곱합은 모형적합 방식에 따른 변동량의 차로 계산되고 이 차의 계산을 위해 반복적으로 모형을 적합시켜야 하는 번거로움이 발생한다. 반복적인 모형적합의 번거로움을 해소하기 위해 사영이 행해지는 사영공간의 사영행렬을 이용할 수 있다. 따라서 이원고정효과 모형의 가정하에 변동요인에 따른 제2종 제곱합을 사영에 의해 구하는 방법을 제공하고자 한다. 4절의 구체적 자료 예에서 사영과 사영행렬 그리고 제2종 제곱합에 해당하는 사영까지의 거리제곱은 모두 프로그램 R로 구해진 값이다. 즉, 모형의 비교방식을 통해 변동요인의 제곱합을 구하는 제2종 분석과는 달리 사영공간으로의 사영행렬을 직접적으로 이용한 점에서 분석방법에 분명한 차이가 있다. 사영에 대한 구체적 논의는 Johnson과 Wichern (1988) 그리고 Graybill (1983) 등에서 살펴볼 수 있다.

2. 모형에 대한 가정

실험자료를 분석하기 위한 모형으로 이원 고정효과 모형을 가정한다. 실험단위의 반응에 영향을 미치는 두개의 처리요인을 각기 A와 B라 두자. 요인 A의 수준 수는 a개이고 요인 B의 수준 수는 b개라 가정한다. 두 고정요인 A와 B의 수준결합으로 주어지는 처리조합의 수는 모두 ab개가 되고, 요인 A의 수준 i와 요인 B의 수준 j에서의 처리조합은 (i, j)로 표시된다. 요인 A의 i번째 수준의 효과를 α_i $(i = 1, 2, \cdots, a)$라 둘 때, α_i는 고정효과를 나타낸다. 요인 B의 j번째 수준의 효과를 β_j $(j = 1, 2, \cdots, b)$라 둘 때, β_j는 고정효과를 나타낸다. 처리조합 (i, j)가 n_{ij}개 실험단위들에 임의로

배정된다고 가정한다. 실험단위들이 거의 동질적이라 가정할 때 ab개의 처리조합은 실험단위들에 임의로 배정된다. 처리조합 (i,j)가 행해진 k번째 실험단위에서의 관측값을 y_{ijk}라 둘 때 고정효과 모형은 다음과 같다.

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \tag{2.1}$$

단, μ는 전체평균이고 α_i와 β_j는 각기 요인 A의 수준 i에서의 고정효과와 B의 수준 j에서의 고정효과를 나타낸다. $(\alpha\beta)_{ij}$는 처리조합 (i,j)에서의 두 요인의 교호작용을 나타낸다. 오차항 ϵ_{ijk}는 평균이 0이고 분산이 σ_ϵ^2이라 가정한다. 모형내 추정되어야 할 미지모수들의 수는 $1+a+b+ab$개이고 추정방법으로 최소제곱법을 이용한다고 가정한다. 최소제곱법으로 주어지는 정규방정식의 수는 추정해야 할 모수의 수와 일치하나 모형내 미지모수들의 해를 얻기 위한 정규방정식들이 종속적이므로 해를 구하기 위한 제약식이 주어진다. 제약식으로 $\sum_{i=1}^a \alpha_i = 0$, $\sum_{j=1}^b \beta_j = 0$이고 $\sum_{i=1}^a (\alpha\beta)_{ij} = \sum_{j=1}^b (\alpha\beta)_{ij} = 0$ 임을 이용하여 정규방정식들에 대한 유일 해를 구할 수 있다. 해를 구한 후 두 요인의 수준변화와 결부된 변동량을 계산하게 된다. 요인 A의 수준변화에 따른 제곱합을 SSA, 요인 B의 수준변화에 따른 제곱합을 SSB, 두 요인의 교호작용에 기인한 제곱합을 $SSAB$라 두자. ab개의 처리조합 (i,j)가 $n = \sum_{i=1}^a \sum_{j=1}^b n_{ij}$에서 측정된 값의 총변동량을 TSS라 두자. 식 (2.1)의 가정하에 처리조합에 따른 제곱합을 SST라 두자. 전체평균 μ에 따른 변동량을 $SS\mu$라 두면 SST는 $SS\mu$, SSA, SSB와 $SSAB$의 네 성분으로 분해될 수 있다. 식 (2.1)의 모형을 모수가 기입된 순서대로 순차적으로 적합시켜 네 성분에 대한 제곱합을 구하게 되면 SST는 네 성분들의 합으로 표현됨을 알 수 있다. 이 경우의 제곱합을 제1종 분석에 의한 제1종 제곱합이라 부른다. 자료가 균형자료인 경우에 모형의 다른 적합방식에 대해서도 동일한 결과를 얻을 수 있으나 불균형 자료의 경우 이러한 결과는 일반적으로 성립하지 않게 된다.

제2종 제곱합은 가정된 모형의 순차적 적합이 아닌 모형비교 방식에 의해 계산된다.균형자료인 경우 제1종 분석에서와 같이 동일한 결과를 얻을 수 있으나 불균형 자료의 경우 SST는 네 성분들의 합으로 표현되지 않는다. SST가 모형에서 고려된 성분들의 제곱합으로 분해되지 않는 논리적 근거를 사영의 개념에 기초하여 마련하기 위해 식 (2.1)을 다음과 같이 행렬로 표현한다.

$$\boldsymbol{y} = \boldsymbol{j}\mu + \boldsymbol{X}_A\boldsymbol{\alpha} + \boldsymbol{X}_B\boldsymbol{\beta} + \boldsymbol{X}_{AB}(\boldsymbol{\alpha\beta}) + \boldsymbol{\epsilon} \tag{2.2}$$

단, $\boldsymbol{y}$는 $n\times 1$인 관측벡터이고 $n = \sum_{i=1}^a \sum_{j=1}^b n_{ij}$이다. $\boldsymbol{j}$는 모든 원소가 1인 $n\times 1$벡터이다. $\boldsymbol{X}_A$는 크기가 $n\times a$인 계수행렬, $\boldsymbol{X}_B$는 크기가 $n\times b$인 계수행렬이고 $\boldsymbol{X}_{AB}$는 크기가 $n\times ab$인 계수행렬을 나타낸다. $\boldsymbol{\epsilon}$은 $n\times 1$인 오차벡터이고 다변량 정규분포 $N(\boldsymbol{0}, \sigma_\epsilon^2\boldsymbol{I}_n)$를 따른다고 가정한다. $\boldsymbol{\alpha}$는 $a\times 1$인 모수벡터이고, $\boldsymbol{\beta}$는 $b\times 1$인 모수벡터이다. $(\boldsymbol{\alpha\beta})$는 교호작용을 나타내는 벡터이고 크기가 $ab\times 1$이다.

3. 사영을 이용한 제2종 제곱합

행렬모형식 (2.2)로부터 사영을 이용한 제2종 제곱합을 구해 보기로 한다. 식 (2.1)의 가정하에 모형내 모수를 추정하기 위한 최소제곱법의 적용은 행렬로 표현된 식 (2.2)에서는 다차원 공간상에서의 사영과 일치하게 된다. 즉, 크기가 $n\times 1$인 관측벡터 $\boldsymbol{y}$를 n차원의 벡터공간에서 한 벡터로 간주할 때 모형행렬 $\boldsymbol{X} = (\boldsymbol{j} \;\; \boldsymbol{X}_A \;\; \boldsymbol{X}_B \;\; \boldsymbol{X}_{AB})$로생성되는 부분 벡터공간으로의 사영이 됨을 알 수 있다.

제2종 제곱합의 계산은 모형비교 방식을 통해 주어지는 잔차제곱합의 차를 이용하여 구하게 된다. 모형비교 방식에 따른 각 효과의 제곱합은 모형내 그 효과와 동일한 수준에 있거나 하위수준에 있는

모든 다른 효과에 적합된 모형을 비교함으로 구해진다. 모형비교 방식을 사영의 관점에서 다루기 위해 행렬표현식 (2.2)를 이용하기로 한다.

평균 μ에 따른 변동량은 먼저 $y_{ijk} = \mu + \epsilon_{ijk}$를 자료에 적합시켜 최소제곱법에 의해 $\widehat{\mu} = \bar{y}_{\cdots} = (1/n)\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}} y_{ijk}$를 구한다. $\hat{\mu}$을 이용하여 제곱합 $SSM = n\bar{y}_{\cdots}^2$ 을 계산한다.

행렬표현의 모형식 $\boldsymbol{y} = \boldsymbol{j}\mu + \boldsymbol{\epsilon}$을 이용할 때 $\widehat{\mu}$은 관측벡터 $\boldsymbol{y}$를 $\boldsymbol{j}$에 의해 생성된 부분 벡터공간으로의 사영이 행해지는 위치를 나타내며 $\widehat{\mu} = \boldsymbol{j}^{-}\boldsymbol{y}$로 주어진다. 단, $\boldsymbol{j}^{-} = (\boldsymbol{j}'\boldsymbol{j})^{-1}\boldsymbol{j}'$이다. 관측벡터 $\boldsymbol{y}$의 $\boldsymbol{j}$로의 사영에 의한 제곱거리는 $\boldsymbol{y}'\boldsymbol{j}\boldsymbol{j}^{-}\boldsymbol{y}$가 되고 SSM과 일치하게 된다. 즉, $SSM = \boldsymbol{y}'\boldsymbol{j}\boldsymbol{j}^{-}\boldsymbol{y}$이다.

요인 A의 수준별 효과에 따르는 제2종 제곱합을 $R(\alpha|\mu,\beta)$라 두자. $R(\alpha|\mu,\beta)$의 의미는 μ와 β다음에 α를 적합시키는 데 따르는 제곱합에서의 감소량을 나타낸다. 이 감소량은 $R(\alpha|\mu,\beta) = RSS(\mu,\beta) - RSS(\mu,\beta,\alpha)$에 의해 구해진다. 여기서 $RSS(\cdot)$는 잔차제곱합을 의미한다. 즉, $RSS(\mu,\beta)$는 모형 $y_{ijk} = \mu + \beta_j + \epsilon_{ijk}$를 자료에 적합시켰을 때 주어지는 잔차제곱합이다. 행렬모형식으로 살펴보면 $\boldsymbol{y} = \boldsymbol{j}\mu + \boldsymbol{X}_B\boldsymbol{\beta} + \boldsymbol{\epsilon}$가 되고 이는 관측벡타 $\boldsymbol{y}$의 $\boldsymbol{X}_{IB} = (\boldsymbol{j} \quad \boldsymbol{X}_B)$로의 사영을 의미한다. 사영으로부터 주어지는 잔차제곱합 $RSS(\mu,\boldsymbol{\beta}) = \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{X}_{IB}\boldsymbol{X}_{IB}^{-})\boldsymbol{y}$이다. 잔차제곱합 $RSS(\mu,\beta,\alpha)$는 모형 $y_{ijk} = \mu + \beta_j + \alpha_i + \epsilon_{ijk}$를 자료에 적합시켰을 때 잔차로부터 계산되는 양이다. 모형 $y_{ijk} = \mu + \beta_j + \alpha_i + \epsilon_{ijk}$를 자료에 적합시키는 부분은 모형식 $\boldsymbol{y} = \boldsymbol{j}\mu + \boldsymbol{X}_B\boldsymbol{\beta} + \boldsymbol{X}_A\boldsymbol{\alpha} + \boldsymbol{\epsilon}$에서 관측벡터 $\boldsymbol{y}$의 $\boldsymbol{X}_{IBA} = (\boldsymbol{j} \quad \boldsymbol{X}_B \quad \boldsymbol{X}_A)$로의 사영에 해당한다. 사영이 행해진 후 잔차제곱합 $RSS(\mu,\beta,\alpha)$에 해당하는 양은 $\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{X}_{IBA}\boldsymbol{X}_{IBA}^{-})\boldsymbol{y}$가 되고 $RSS(\mu,\boldsymbol{\beta},\boldsymbol{\alpha})$로 나타낼 수 있다. 행렬로 표현할 때 모수효과 벡터 $\boldsymbol{\alpha}$에 기인한 제2종 제곱합을 $R(\boldsymbol{\alpha}|\mu,\boldsymbol{\beta})$로 나타낸다.

$$\begin{aligned} R(\boldsymbol{\alpha}|\mu,\boldsymbol{\beta}) &= RSS(\mu,\boldsymbol{\beta}) - RSS(\mu,\boldsymbol{\beta},\boldsymbol{\alpha}) \\ &= \boldsymbol{y}'(\boldsymbol{X}_{IBA}\boldsymbol{X}_{IBA}^{-} - \boldsymbol{X}_{IB}\boldsymbol{X}_{IB}^{-})\boldsymbol{y} \end{aligned} \tag{3.1}$$

식 (3.1)의 첫 번째 식에서 모수효과 벡터 $\boldsymbol{\alpha}$에 따른 제2종 제곱합은 관측벡터 $\boldsymbol{y}$의 $\boldsymbol{X}_{IB}$로의 사영에 의한 잔차제곱합과 $\boldsymbol{\alpha}$가 포함된 모수벡터 $(\mu\ \boldsymbol{\beta}'\ \boldsymbol{\alpha}')'$의 계수행렬 $\boldsymbol{X}_{IBA}$로의 사영에서 주어지는 잔차제곱합 간의 차이로 계산된다. 식 (3.1)의 두 번째 식은 두 잔차제곱합 간의 차는 결과적으로 관측벡터 $\boldsymbol{y}$의 $\boldsymbol{X}_{IBA}$로의 사영과 $\boldsymbol{X}_{IB}$로의 사영에서 각 사영까지의 거리제곱 간의 차이임을 나타내고 있다.

요인 B의 수준별 효과에 기인하는 제2종 제곱합을 $R(\beta|\mu,\alpha)$라 두자. $R(\beta|\mu,\alpha)$의 의미는 μ와 α다음에 β를 적합시키는 데 따르는 제곱합에서의 감소량을 나타낸다. 이 감소량은 $R(\beta|\mu,\alpha) = RSS(\mu,\alpha) - RSS(\mu,\alpha,\beta)$에 의해 구해진다. $RSS(\mu,\alpha)$를 구하기 위한 행렬식을 살펴보면 $\boldsymbol{y} = \boldsymbol{j}\mu + \boldsymbol{X}_A\boldsymbol{\alpha} + \boldsymbol{\epsilon}$가 되고 이는 관측벡터 $\boldsymbol{y}$의 $\boldsymbol{X}_{IA} = (\boldsymbol{j} \quad \boldsymbol{X}_A)$로 생성된 공간으로의 사영을 의미한다. 사영으로부터 주어지는 잔차제곱합을 $RSS(\mu,\boldsymbol{\alpha})$라 둘 때, $RSS(\mu,\boldsymbol{\alpha}) = \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{X}_{IA}\boldsymbol{X}_{IA}^{-})\boldsymbol{y}$이다. 잔차제곱합 $RSS(\mu,\alpha,\beta)$는 모형 $y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$를 자료에 적합시켰을 때 잔차로부터 계산되는 양이다. 모형 $y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$를 자료에 적합시키는 부분은 모형식 $\boldsymbol{y} = \boldsymbol{j}\mu + \boldsymbol{X}_A\boldsymbol{\alpha} + \boldsymbol{X}_B\boldsymbol{\beta} + \boldsymbol{\epsilon}$에서 관측벡터 $\boldsymbol{y}$의 $\boldsymbol{X}_{IAB} = (\boldsymbol{j} \quad \boldsymbol{X}_A \quad \boldsymbol{X}_B)$로의 사영에 해당한다. 사영이 행해진 후 잔차제곱합 $RSS(\mu,\alpha,\beta)$에 해당하는 양은 $\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{X}_{IAB}\boldsymbol{X}_{IAB}^{-})\boldsymbol{y}$가 되고 $RSS(\mu,\boldsymbol{\alpha},\boldsymbol{\beta})$로 나타낼 수 있다. 행렬로 표현할 때 모수효과 벡터 $\boldsymbol{\beta}$에 기인한 제2종 제곱합은 $R(\boldsymbol{\beta}|\mu,\boldsymbol{\alpha})$로 나타낸다.

$$\begin{aligned} R(\boldsymbol{\beta}|\mu,\boldsymbol{\alpha}) &= RSS(\mu,\boldsymbol{\alpha}) - RSS(\mu,\boldsymbol{\alpha},\boldsymbol{\beta}) \\ &= \boldsymbol{y}'(\boldsymbol{X}_{IAB}\boldsymbol{X}_{IAB}^{-} - \boldsymbol{X}_{IA}\boldsymbol{X}_{IA}^{-})\boldsymbol{y} \end{aligned} \tag{3.2}$$

식 (3.2)의 처음 식은 모수효과 벡터 $\boldsymbol{\beta}$에 따른 제2종 제곱합의 계산을 위해 모형의 적합에 따른 두 잔차제곱합을 구해야 함을 보여주고 있다. 그러나 식 (3.2)의 두 번째는 관측벡터 $\boldsymbol{y}$의 $\boldsymbol{X}_{IAB}$로의 사영과 $\boldsymbol{X}_{IA}$로의 사영이 행해졌을 때 각 사영까지의 거리제곱 간의 차이로 구해질 수 있음을 보여준다.

교호작용 $(\alpha\beta)$효과에 따른 제2종 제곱합을 $R((\alpha\beta)|\mu,\alpha,\beta)$라 두자. $R((\alpha\beta)|\mu,\alpha,\beta)$의 의미는 μ, α, β의 적합 후에 $(\alpha\beta)$를 적합시킬 때 교호작용에 따르는 제곱합에서의 감소량을 나타낸다. 이 감소량은 $R((\alpha\beta)|\mu,\alpha,\beta)= RSS(\mu,\alpha,\beta) - RSS(\mu,\alpha,\beta,(\alpha\beta))$에 의해 구해진다. $RSS(\mu,\alpha,\beta)$에 해당하는 $RSS(\mu,\boldsymbol{\alpha},\boldsymbol{\beta})$를 행렬표현식으로부터 구하였으므로 $RSS(\mu,\alpha,\beta,(\alpha\beta))$를 생각해 보자. 잔차제곱합 $RSS(\mu,\alpha,\beta,(\alpha\beta))$는 모형 $y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$를 자료에 적합시켰을 때 잔차로부터 계산되는 양이다. 모형 $y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$를 자료에 적합시키는 부분은 행렬모형식 (2.2)에서 관측벡터 $\boldsymbol{y}$의 모형행렬 $\boldsymbol{X}$에 의해 생성되는 공간으로의 사영에 해당한다. 사영이 행해진 후 잔차제곱합 $RSS(\mu,\alpha,\beta,(\alpha\beta))$에 해당하는 양은 $\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{y}$가 되므로 $RSS(\mu,\boldsymbol{\alpha},\boldsymbol{\beta},(\boldsymbol{\alpha\beta}))$로 나타낼 수 있다. 행렬로 표현할 때 모수효과 벡타 $(\boldsymbol{\alpha\beta})$에 기인한 제2종 제곱합을 $R((\boldsymbol{\alpha\beta})|\mu,\boldsymbol{\alpha},\boldsymbol{\beta})$라 두면

$$\begin{aligned} R((\boldsymbol{\alpha\beta})|\mu,\boldsymbol{\alpha},\boldsymbol{\beta}) &= RSS(\mu,\boldsymbol{\alpha},\boldsymbol{\beta}) - RSS(\mu,\boldsymbol{\alpha},\boldsymbol{\beta},(\boldsymbol{\alpha\beta})) \\ &= \boldsymbol{y}'(\boldsymbol{X}\boldsymbol{X}^- - \boldsymbol{X}_{IAB}\boldsymbol{X}_{IAB}^-)\boldsymbol{y} \end{aligned} \tag{3.3}$$

로 계산된다. 식 (3.3)의 첫 식은 관측벡터 $\boldsymbol{y}$의 $\boldsymbol{X}_{IAB}$로의 사영에서 주어지는 잔차제곱합과 모형행렬 $\boldsymbol{X}$에 의해 생성되는 부분공간으로의 사영에서 주어지는 잔차제곱합 간의 차이로 계산됨을 나타낸다. 식 (3.3)의 둘째 식은 관측벡터 $\boldsymbol{y}$의 모형행렬 $\boldsymbol{X}$로의 사영에서 주어지는 거리제곱과 $\boldsymbol{X}_{IAB}$로의 사영에서 주어지는 거리제곱간의 차이로 계산됨을 보여준다.고정효과에 따른 제2종 제곱합을 구하기 위한 모형비교의 적합방식에서 관련된 모형행렬들을 나타내기 위해 식 (2.2)를

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\epsilon} \tag{3.4}$$

로 표현한다. 여기서 $\boldsymbol{\theta'}$는 $(\mu\ \ \boldsymbol{\alpha}'\ \ \boldsymbol{\beta}'\ \ (\boldsymbol{\alpha\beta})')$인 모수벡터이다. 모형행렬 $\boldsymbol{X}$에 의해 생성된 벡터공간을 $\mathfrak{X}$라 둘 때, 벡터공간 $\mathfrak{X}$는 모수들의 추정공간이 된다. 최소제곱법에 의한 모수추정은 관측벡터 $\boldsymbol{y}$의 벡터공간 $\mathfrak{X}$내 부분 벡터공간으로의 사영에 의한 추정을 의미하게 된다. 사영에 의한 모수추정에서 모수추정 방법은 모수 또는 모수벡터가 기입된 순서대로 관련된 계수행렬로 생성된 벡터공간 $\mathfrak{X}$의 부분공간으로의 사영을 이용할 수 있다. 이와 관련하여 Choi (2010)은 사영을 이용한 순차적 분석을 다루고 있다.

관심의 모수효과에 따른 제2종 제곱합은 동일 수준 또는 하위수준의 모수벡터와 관련된 모수공간으로의 사영에 따른 잔차제곱합과 추정하고자하는 모수벡터가 추가로 포함된 모수공간으로의 사영에서 주어진 잔차제곱합과의 차이로 얻게 된다. 이 경우에 관심효과에 따른 제곱합을 구하기 위해 잔차제곱합이 반복되어 계산되어야 하는 번거로움이 있다. 이러한 번거로움을 해소하기 위해 사영행렬을 이용할 수 있음을 논의한다.

균형자료의 경우 요인별 효과에 따른 제2종 제곱합은 모형의 순차적 분석에서 구해지는 제1종 제곱합과 동일하다. 균형자료의 경우 모수추정이 행해지는 부분공간들이 상호 직교하므로 모형의 적합방식에 상관없이 모수에 따른 제곱합은 일정하게 구해진다.

불균형 자료의 분석에서 요인 A의 효과벡터 $\boldsymbol{\alpha}$에 따른 제2종 제곱합을 모형행렬 $\boldsymbol{X}$에 의해 생성된 벡터공간 $\mathfrak{X}$로의 사영으로 구해본다. 모형비교를 이용한 제2종 제곱합은 사영공간인 $\mathfrak{X}$를 몇 개의 부분공간으로 구분하여 각 모수벡터와 관련된 부분공간에서의 사영을 이용하게 된다. 개별 모수벡터에 따른 제2종 제곱합은 해당하는 부분공간으로의 사영행렬로 표시되는 2차형식으로 주어진다. 모수벡터 $\boldsymbol{\alpha}$에 따른 제2종 제곱합을 SSA라 두자. SSA를 계산하기 위한 사영공간은 모수벡터 $\boldsymbol{\alpha}$와 동일한 수준의 모수벡터 $\boldsymbol{\beta}$와 하위수준의 μ가 포함된 계수행렬 $\boldsymbol{X}_{IAB} = (\boldsymbol{j}\ \ \boldsymbol{X}_A\ \ \boldsymbol{X}_B)$에 의해 생성되는 부분공간이다. SSA를 계산하기 위한 과정으로 관측벡터 $\boldsymbol{y}$의 $\boldsymbol{X}_{IB}$로의 사영은 모형식 (3.4)

에서 축소된 모형

$$\boldsymbol{y} = \boldsymbol{X}_{IAB}\boldsymbol{\theta}_{IAB} + \boldsymbol{\epsilon} \tag{3.5}$$

의 적합을 의미한다. 단, $\boldsymbol{\theta}_{IAB}$는 $(\mu\ \boldsymbol{\alpha}'\ \boldsymbol{\beta}')'$를 나타낸다. 이때의 사영은 $\boldsymbol{X}_{IAB}\boldsymbol{X}_{IAB}^{-}\boldsymbol{y}$가 된다. 사영까지의 거리제곱은 $\boldsymbol{y}'\boldsymbol{X}_{IAB}\boldsymbol{X}_{IAB}^{-}\boldsymbol{y}$이다. 관측벡터 $\boldsymbol{y}$의 $\boldsymbol{X}_{IB}$로의 사영에 해당하는 모형은

$$\boldsymbol{y} = \boldsymbol{X}_{IB}\boldsymbol{\theta}_{IB} + \boldsymbol{\epsilon} \tag{3.6}$$

이다. 이때의 사영은 $\boldsymbol{X}_{IB}\boldsymbol{X}_{IB}^{-}\boldsymbol{y}$가 된다. 사영까지의 거리제곱은 $\boldsymbol{y}'\boldsymbol{X}_{IB}\boldsymbol{X}_{IB}^{-}\boldsymbol{y}$이다. 두 사영까지의 거리제곱 간의 차이는 $\boldsymbol{y}'\boldsymbol{X}_{IAB}\boldsymbol{X}_{IAB}^{-}\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}_{IB}\boldsymbol{X}_{IB}^{-}\boldsymbol{y} = \boldsymbol{y}'(\boldsymbol{X}_{IAB}\boldsymbol{X}_{IAB}^{-} - \boldsymbol{X}_{IB}\boldsymbol{X}_{IB}^{-})\boldsymbol{y}$이다. 즉, $SSA = \boldsymbol{y}'(\boldsymbol{X}_{IAB}\boldsymbol{X}_{IAB}^{-} - \boldsymbol{X}_{IB}\boldsymbol{X}_{IB}^{-})\boldsymbol{y}$이고 식 (3.1)의 $R(\boldsymbol{\alpha}|\mu,\boldsymbol{\beta})$와 동일하다. SSA는 모수벡터의 추정공간으로의 사영으로부터 계산되었고 $R(\boldsymbol{\alpha}|\mu,\boldsymbol{\beta})$는 모형의 적합에서 계산된 잔차제곱합의 차이로 얻어진 양이라는 점에 차이가 있다.

모수벡터 $\boldsymbol{\beta}$에 따른 제2종 제곱합을 SSB라 두자. SSB를 계산하기 위한 사영공간은 모수벡터 $\boldsymbol{\beta}$와 동일한 수준의 모수벡터 $\boldsymbol{\alpha}$와 하위수준의 μ가 포함된 계수행렬 $\boldsymbol{X}_{IAB}$에 의해 생성되는 부분공간이다. SSB를 계산하기 위한 사영공간은 SSA를 계산하기 위한 사영공간과 동일하게 되고 관측벡터 $\boldsymbol{y}$의 $\boldsymbol{X}_{IA}$로의 사영에 해당하는 모형은

$$\boldsymbol{y} = \boldsymbol{X}_{IA}\boldsymbol{\theta}_{IA} + \boldsymbol{\epsilon} \tag{3.7}$$

이다. 단, $\boldsymbol{\theta}_{IA}' = (\mu\ \boldsymbol{\alpha}')'$이다. 이때의 사영은 $\boldsymbol{X}_{IA}\boldsymbol{X}_{IA}^{-}\boldsymbol{y}$가 된다. 사영까지의 거리제곱은 $\boldsymbol{y}'\boldsymbol{X}_{IA}\boldsymbol{X}_{IA}^{-}\boldsymbol{y}$이다. 두 사영까지의 거리제곱 간의 차이는 $SSB = \boldsymbol{y}'(\boldsymbol{X}_{IAB}\boldsymbol{X}_{IAB}^{-} - \boldsymbol{X}_{IA}\boldsymbol{X}_{IA}^{-})\boldsymbol{y}$이고 식 (3.2)의 $R(\boldsymbol{\beta}|\mu,\boldsymbol{\alpha})$와 동일하다.

요인 A와 B의 교호작용 벡터를 나타내는 $(\boldsymbol{\alpha\beta})$에 따른 제2종 제곱합을 $SSAB$라 두자. $SSAB$를 계산하기 위한 사영공간은 교호작용 벡터 $(\boldsymbol{\alpha\beta})$와 동일한 수준의 모수벡터 또는 하위수준의 모수가 포함된 계수행렬에 의해 생성된 벡터공간은 모형행렬 $\boldsymbol{X}$에 의해 생성된 벡터공간 $\mathfrak{X}$가 된다. 관측벡터 $\boldsymbol{y}$의 $\boldsymbol{X}$로의 사영에 대한 모형으로 모형식 (3.4)를 적합시킨다. 이때의 사영은 $\boldsymbol{X}\boldsymbol{X}^{-}\boldsymbol{y}$가 된다. 사영까지의 제곱거리는 $\boldsymbol{y}'\boldsymbol{X}\boldsymbol{X}^{-}\boldsymbol{y}$이다. 교호작용 벡터 $(\boldsymbol{\alpha\beta})$를 포함하고 있지 않은 사영공간은 $\boldsymbol{X}_{IAB}$에 의해 생성되는 부분공간이다. 이 공간으로의 사영에 해당하는 거리제곱은 $\boldsymbol{y}'\boldsymbol{X}_{IAB}\boldsymbol{X}_{IAB}^{-}\boldsymbol{y}$이다. 따라서 $SSAB = \boldsymbol{y}'(\boldsymbol{X}\boldsymbol{X}^{-} - \boldsymbol{X}_{IAB}\boldsymbol{X}_{IAB}^{-})\boldsymbol{y}$이고 식 (3.3)의 $R((\boldsymbol{\alpha\beta})|\mu,\boldsymbol{\alpha},\boldsymbol{\beta})$에 해당한다.

모수벡터 $\boldsymbol{\alpha}$에 따른 제2종 제곱합을 구하기 위한 사영공간과 $\boldsymbol{\beta}$에 따른 제2종 제곱합을 구하기 위한 사영공간은 부분행렬 $\boldsymbol{X}_{IAB} = (\boldsymbol{j}\quad \boldsymbol{X}_A\quad \boldsymbol{X}_B)$에 의해 동일하게 주어지는 공간이다. 그러나 불균형자료의 경우에는 각 모수벡터와 관련된 부분공간들은 상호 직교하는 부분공간들로 구성되지 않게 된다.

4. 자료의 예

행렬모형식 (3.1)의 가정하에 사영을 이용하여 분석할 자료로 Milliken과 Johnson (1984)에서 사용된 불균형 자료를 자료분석의 예로 사용한다. Figure 4.1은 완전임의 설계구조 (completely randomized design structure)에서 행해진 이원 처리구조의 실험으로부터 주어진 자료를 나타낸다. 표의 자료에서부터 요인 A는 A_1, A_2의 두 수준 이고 요인 B는 B_1, B_2, B_3의 세 수준이다. 따라서 처리조합은 이원구조의 6개로 두어진다. 실험에 이용될 실험단위들이 동질적이라 간주될 때 실험의 설계구조는 완전임의 설계구조가 된다.

Table 4.1 Unbalanced two-way experimental data

요인 A	요인 B		
	B_1	B_2	B_3
A_1	19, 20, 21	24, 26	22, 25, 25
A_2	25, 27	21, 24, 24	31, 32, 33
합	112	119	168

위자료는 각 처리조합에서의 관측수가 동일하지 않으므로 불균형 자료로 분류된다. 자료의 행렬표현으로부터 관측벡터 $\boldsymbol{y}$는 크기가 16×1인 열벡터이다. 모형행렬 $\boldsymbol{X}$로의 사영은 $\boldsymbol{X}\boldsymbol{X}^{-}\boldsymbol{y}$이고 사영까지의 거리제곱은 $\boldsymbol{y}'\boldsymbol{X}\boldsymbol{X}^{-}\boldsymbol{y}$의 2차 형식으로 표현되며 이는 처리조합에 따른 제곱합 SST이다. $\boldsymbol{y}'\boldsymbol{X}\boldsymbol{X}^{-}\boldsymbol{y}$=10,189로 계산된다. 평균 μ에 따른 제곱합 SSM은 $\boldsymbol{y}'\boldsymbol{j}\boldsymbol{j}^{-}\boldsymbol{y}$=9950.062이므로 $\boldsymbol{y}'(\boldsymbol{X}\boldsymbol{X}^{-} - \boldsymbol{j}\boldsymbol{j}^{-})\boldsymbol{y}$=238.938이다. 이는 모수추정 공간에서 μ와 관련된 벡터 $\boldsymbol{j}$로의 사영이 이루어지고 그 사영까지의 제곱거리를 제외한 양이다. 모수효과 벡터 $\boldsymbol{\alpha}$에 따른 제2종 제곱합 SSA를 계산하기 위한 사영공간은 두 개의 부분공간으로 나누어진다. 즉, $\boldsymbol{X}_{IAB} = (\boldsymbol{j} \quad \boldsymbol{X}_A \quad \boldsymbol{X}_B)$로 생성되는 사영공간은 $\boldsymbol{X}_{IB} = (\boldsymbol{j} \quad \boldsymbol{X}_B)$에 의한 부분공간과 이와 직교하는 관심의 두 부분공간으로 나누어지므로 요인 A의 모수효과 벡터 $\boldsymbol{\alpha}$에 따른 변동량을 계산할 수 있다. 관측벡터 $\boldsymbol{y}$의 2차형식 $\boldsymbol{y}'(\boldsymbol{X}_{IAB}\boldsymbol{X}_{IAB}^{-} - \boldsymbol{X}_{IB}\boldsymbol{X}_{IB}^{-})\boldsymbol{y}$에 의해 구해진 SSA=72.369로 구해진다. 요인 B의 효과벡터 $\boldsymbol{\beta}$에 따른 제2종 제곱합 SSB= $\boldsymbol{y}'(\boldsymbol{X}_{IAB}\boldsymbol{X}_{IAB}^{-} - \boldsymbol{X}_{IA}\boldsymbol{X}_{IA}^{-})\boldsymbol{y}$는 90.744로 구해진다. 요인 A와 B의 교호작용 벡터를 나타내는 $(\boldsymbol{\alpha\beta})$에 따른 제2종 제곱합 $SSAB$는 2차형식 $\boldsymbol{y}'(\boldsymbol{X}\boldsymbol{X}^{-} - \boldsymbol{X}_{IAB}\boldsymbol{X}_{IAB}^{-})\boldsymbol{y}$로 계산되며 그 값은 71.631이다.

불균형 자료에서는 변동요인에 따른 제2종 제곱합의 총합은 처리조합에 따른 제곱합과 일치하지 않음을 보여 주고 있다. 순차적 분석에 의한 요인들의 제1종 제곱합을 살펴보면 SSA와 $SSAB$는 동일하나 요인 B의 효과벡터 $\boldsymbol{\beta}$에 따른 제2종 제곱합 SSB가 제1종 제곱합의 경우 94.938로 주어진다. 이는 변동량에 있어 4.194의 차이가 있음을 보여 준다. 이러한 차이는 사영이 행해지는 사영공간의 분할에 있어 $\boldsymbol{\alpha}$와 관련된 부분공간과 $\boldsymbol{\beta}$가 관련된 부분공간이 상충되기 때문이다. 즉, $\boldsymbol{\alpha}$의 추정에 이용된 사영공간의 고유벡터는 $\boldsymbol{\beta}$의 추정에 이용되는 사영공간의 고유벡터와 직교하지 않는 데 있다. 요인 B의 효과벡터 $\boldsymbol{\beta}$에 따른 제2종 제곱합의 차이 4.194의 값을 계산하는 데 이용된 2차형식들은 $\boldsymbol{y}'[((\boldsymbol{X}_{IAB}\boldsymbol{X}_{IAB}^{-} - \boldsymbol{X}_{IB}\boldsymbol{X}_{IB}^{-})\boldsymbol{X}_A][(\boldsymbol{X}_{IAB}\boldsymbol{X}_{IAB}^{-} - \boldsymbol{X}_{IB}\boldsymbol{X}_{IB}^{-})\boldsymbol{X}_A]^{-}\boldsymbol{y}$와 SSB간의 차로 구해진다.

5. 결론

본 논문은 실험자료의 분석모형으로 이원 고정효과 모형을 가정하고 있다. 실험에 이용되는 실험단위들은 거의 모두 동질적 이라는 가정하에 실험 설계구조로 완전임의배열법을 가정한다. 처리조합을 나타내는 처리구조는 두 요인의 수준결합으로 구성되며 실험단위의 반응값은 요인들의 주효과와 교호작용 그리고 오차를 반영하는 선형모형하에서 분석된다. 일차원상의 관측자료를 행렬로 표현하여 다차원상의 벡터로 간주할 때 선형모형 또한 다차원상의 공간벡터에 대한 행렬모형식에 의해 표현된다. 일차원상의 관측자료를 선형모형의 가정하에 최소제곱법으로 모형식을 추정하는 방법은 다차원상의 공간벡터를 모형행렬로 주어지는 벡터공간으로의 사영에 해당한다는 점에 착안하고 있다. 따라서, 고정효과에 따른 변동량을 구하기 위한 방법으로 사영을 이용하여 관측벡터 $\boldsymbol{y}$의 이차형식으로 구하는 방법을 논의하고 있다. 전통적인 ANOVA 분석방법에서 이용되는 제2종 분석과 관련된 기법이 다차원상의 벡터공간에서는 어떻게 적용되어야 하는 가를 구체적으로 다루고 있다. 또한, 불균형 자료의 분석에서 제2종 분석과 제1종 분석의 차이점을 규명하고 그 차이가 어디에서

발생하고 있는 가를 사영의 측면에서 살펴볼 수 있음을 보여 주고 있다.

참고문헌

Choi, J. S. (2010). A mixed model for repeated split-plot data. *Journal of the Korean Data & Information Science Society*, **21**, 1-9.

Choi, J. S. (2011). Type I analysis by projections, *The Korean Journal of Applied Statistics*, 24, 373-381.

Galecki, A. T. (1994). General class of covariance structures for two or more repeated factors in longitudinal data analysis . *Communications in Statistics-Theory and Methods*, **23**, 3105-3119.

Graybill, F. A. (1976). *Theory and application of the linear model*, Wadsworth Publishing Company, Inc., Belmont.

Graybill, F. A. (1983). *Matrices with applications in statistics*, Wadsworth Publishing Company, Inc., Belmont.

Johnson, R. A. and Wichern, D. W. (1988). *Applied multivariate statistical analysis,* 2nd edition, Prentice Hall, Inc., Englewood Cliffs.

Milliken, G. A. and Johnson, D. E. (1984). *Analysis of messy data*, Van Nostrand Reinhold, New York.

Searle, S. R. (1971). *Linear models*, John Wiley and Sons, Inc., New York.

Journal of the Korean Data & Information Science Society
2012, **23**(6), 1155–1163

http://dx.doi.org/10.7465/jkdi.2012.23.6.1155
한국데이터정보과학회지

Type II analysis by projections†

Jaesung Choi[1]

[1]Department of Statistics, Keimyung University

Received 17 October 2012, revised 11 November 2012, accepted 17 November 2012

Abstract

This paper suggests a method for getting sums of squares due to sources of variation under the assumption of two-way fixed effects model. The method used for calculating the quantities due to fixed-effects is based on the projections of an observation vector $\boldsymbol{y}$ on the column space generated by the model matrix $\boldsymbol{X}$ under the assumed model. The suggested method shows that the calculation of Type II sums of squares by projections is much easier than the classical Type II analysis.

Keywords: Fixed effects model, projection, Type II analysis, unbalanced data.

† The research was supported by the Bisa Research Grant of Keimyung University in 2012.

1 Professor, Department of Statistics, Keimyung University, Daegu 704-701, Korea.
E-mail: jschoi@kmu.ac.kr